HOW ACCURATE ARE INTEREST INVENTORIES? A QUANTITATIVE REVIEW OF CAREER CHOICE HIT RATES

BY

ALEXIS VICTORIA HANNA

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Master's Committee:

Professor James Rounds, Chair
Professor Daniel A. Newman

ABSTRACT

Every year, millions of people use interest inventories to help them make educational and career choices. The present meta-analysis examines the criterion-related validity of interest inventories in predicting career choice outcomes, such as college major choice and occupational membership. This analysis of predictive hit rates incorporates over 75 years of research investigating the accuracy of interest inventories. Using a binomial-normal meta-analytic model to quantitatively estimate the overall hit rate, the present analysis found that measured interests attain an estimated predictive accuracy rate of 50.3% in successfully predicting career choice. Due to a substantial amount of true heterogeneity in effect sizes, we tested several potential moderators. In particular, the hit rate accuracy was moderated by concurrent versus predictive criterion assessment, criterion interest category, interest inventory, type of interest inventory scale, career choice criterion, method used to match criterion to a scale, and hit rate method. Additionally, the present study is an attempt to reintroduce base rates into the evaluation of predictive accuracy. Implications for future research are discussed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Students and adults often turn to interest inventories to help them make one of the most important decisions in their adult life: career choice. The importance of this decision is evidenced by the fact that millions of interest inventories are taken every year in the United States (ACT, 2009; Hansen, 1994; Morris, 2016). Interest assessment is a substantial industry that permeates the fields of education, counseling psychology, organizational psychology, and national testing. Interest inventories attempt to match an individual's interests to those of various occupational groups (Anastasi, 1988; Prediger, 1977; Strong, 1943). Though many researchers have assessed and reviewed the criterion-related validity of these inventories (e.g., Campbell, 1971; Clark, 1961; Kuder, 1977; McArthur, 1954; Spokane, 1979; Strong, 1943; Zytowski, 1976), there has yet to be a quantitative review of these results.

Several prominent interest researchers have provided estimates regarding the accuracy of inventories, but the estimates differ substantially. One of the earliest statements regarding the general accuracy of inventories came from John Holland (1973), who claimed that inventory predictions are only moderately efficient. In the 1980's, two camps emerged. The more pessimistic camp regarded the validity of inventories as "mediocre" (e.g., Crites, 1984, p. 284), whereas the more optimistic side acknowledged a "substantial correspondence" between inventory scores and eventual occupational membership (e.g., Anastasi, 1988, p. 57). In the twenty-first century, estimates became more specific. For example, Fouad (1999) asserted that 40% - 60% of individuals enter occupations that are predicted by their inventory scores. At the high end, Sullivan & Hansen (2004) claimed that as many as two-thirds of U.S. employees are in occupations that match their inventoried interests.

Although broad-reaching validity estimates have been made, it is unclear which estimates, if any, are accurate. Recently, Su (2018) has called for a quantitative summary of the existing literature on interest inventory hit rates. In the present paper, I review almost 100 years of research to establish a meta-analytic estimate of the predictive accuracy of interest inventories. I have three main objectives: to evaluate the criterion-related validity of interest inventories, to examine whether certain characteristics moderate the levels of validity, and to reintroduce base rates into the evaluation of predictive accuracy. In particular, I introduce more accurate base rates by which to compare the accuracy rates for Holland's (1997) six interest categories. In this way, I provide a more accurate estimate of how well interest inventories predict college major choice and eventual occupational membership.

Vocational Interest Assessment

An interest reflects liking of an object or activity (Strong, 1943). Interests are unique constructs in that they are contextualized (Rounds & Su, 2014). Vocational interests in particular are contextualized towards different types of occupations. Vocational interests are important predictors of job performance (Nye, Su, Rounds, & Drasgow, 2012; 2017), turnover intentions (Van Iddekinge, Roth, Putka, & Lanivich, 2011), and career success (Su, 2012). Critically, interests also predict educational and career choices (Dolliver & Worthington, 1981; Hansen & Neuman, 1999; Kuder, 1977; Strong, 1943). The focus of the present paper is to derive meta-analytic estimates to quantify these latter relationships.

Depending on the interest inventory, there are different assumptions about how best to relate interests to occupations. According to E. K. Strong (1927, 1943), interests are on a continuum from like to dislike. This continuum indicates that there should be a substantial relationship between inventory scores and eventual occupational membership. In other words, high scores on an interest inventory should reflect occupations that fit an individual's interests. Strong (1943) proposed that men continuing in an occupation will have higher interest scores in that occupation than in others, and men continuing in an occupation will have higher interest scores in that occupation than men in other occupations. In this way, it is proposed that interests are related to and discriminate between occupations.

Similar to Strong (1943), John Holland (1959) proposed a congruence assumption between people and work environments. Holland's (1959, 1997) interest model contains six interest types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional, abbreviated as RIASEC. Realistic interests include hands-on activities, the outdoors, and practical and physical labor. Investigative involves analytics and scientific thinking. Artistic is characterized by creativity and interest in fine arts. Social revolves around helping and working with others. Enterprising involves persuasiveness and leadership. Finally, conventional involves organization and attention to details (Holland, 1959; 1997). Both people and environments can be classified according to these categories, so the congruence assumption poses that individuals are expected to seek out, remain in, and be satisfied in occupations that match their interest type.

Strong's (1927; 1943) propositions and Holland's (1959; 1997) theory lead to two primary methods of interest inventory construction. Inventories such as the Strong Interest Inventory (SII; Donnay, Morris, Schaubhut, & Thompson, 2004; Harmon, DeWitt, Campbell, & Hansen, 1994) and Kuder Occupational Interest Survey (KOIS; Kuder, 1966) use empirical keying to differentiate between occupational groups. Empirical methods determine the items that most differentiate members of an occupational group from respondents in other occupations. These discriminating items form the occupational scale for that group. For example, if accountants in the norm sample endorse, "I enjoy playing an instrument" more than the rest of the norm sample on average, then that item will be keyed on the Accountant scale. Inventories are then able to match an individual with an occupation when that person responds in a similar way to members of that occupation (Donnay et al., 2004).

Through empirical scale development, items may not reflect the occupations to which they are keyed. Conversely, inventories such as the Self-Directed Search (SDS; Holland, Powell, & Fritzsche, 1994), Vocational Preference Inventory (VPI; Holland, 1985), and Revised Unisex Edition of the ACT Inventory (UNIACT-R; Swaney, Lamb, Prediger, & American College Testing Program, 1995) use a deductive approach to determine an individual's interest group. Rather than data-based keying of occupational scales, rational/theoretical scale development involves the composition of items that align with each theoretical interest area (Burisch, 1984). Individuals receive a score for each interest area, and the interest area corresponding to the highest score is considered the individual's high-point code, or primary interest. Depending on the level of specificity, individuals may receive a two-letter or three-letter high-point code corresponding to their second- and third-highest score areas, respectively. Interest congruence is assessed based on the degree of match between an individual's primary interests and those of their occupation.

Criterion-Related Validity

To examine the full range of the criterion-related validity of interest inventories, both concurrent and predictive validity studies are included in the present meta-analysis. To determine the concurrent validity of an interest inventory, the inventory is administered to a group of employees to determine whether the employees score highly on their own occupational scale (e.g. Dik & Hansen, 2004; Dolliver & Worthington, 1981; Donnay & Borgen, 1996). Alternatively, the inventory may be administered to students to compare their measured interests to their current major (e.g. Gasser, Larson, & Borgen, 2007; Hansen & Neuman, 1999; Saladin, 1995; Williams, 1972). By measuring the interests of people who are currently members of certain occupations or majors, this process determines how well the inventory discriminates between criterion groups. Ultimately, the level of concurrent validity indicates the degree of correspondence between individuals' current occupations or majors and their matched inventory scales (Hansen & Neuman, 1999).

In predictive validity studies, the interest inventory is administered prior to collecting the criterion. The goal of these studies is to determine how well the inventory predicts the individual's eventual occupation (e.g. Bartling & Hood, 1981; Hansen & Dik, 2005; McArthur, 1954; Wiggins & Weslander, 1977), college major choice (e.g. Lunneborg, 1993; O'Neil & Magoon, 1977; Prediger & Johnson, 1979), or other criterion, such as vocational aspiration (e.g. Gottfredson & Holland, 1975; Spokane, 1979a). The time between inventory administration and obtainment of the criterion data ranges from a few weeks (Holland, Gottfredson, & Baker, 1990) to 25 years (Zytowski, 1974). Both concurrent and predictive validity studies are important to assess the degree to which interest inventories accurately relate to the criterion.

Hit rates. To quantify the validity of an inventory, researchers typically calculate the hit rate, or the percentage of correct predictions. This calculation requires every respondent's career choice criterion group to be classified according to a scale on the inventory. The classification determines whether the inventory results predict the criterion group. Researchers then derive a hit calculation based on the scoring scheme of a particular inventory.

For example, if an individual's highest score is on the Accountant occupational scale, and the individual later works as an accountant, that is considered a hit. Similarly, if an individual's highest score is in the Social category, and that individual is employed as an elementary school teacher, that is also considered a hit because the occupation of elementary school teacher is classified as Social (Rounds, Armstrong, Liao, Lewis, & Rivkin, 2008). Essentially, the hit rate determines the percent of respondents that would be referred to their criterion group based on their inventory scores. Using these hit rates, I can assess the accuracy of various interest inventories in predicting career choice.

Base rates. When evaluating the accuracy of a measure, it is important to compare the predictive rates to base rates (Bokhari & Hubert, 2015; Meehl & Rosen, 1955; Schmidt, 1974). Base rates describe the frequency of a group or outcome in the population of interest (Meehl & Rosen, 1955; Schmidt, 1974). However, the choice of this chance rate or frequency may come from several possibilities. For example, one may wish to examine the predictive accuracy of the Self Directed Search in predicting college major choice for females. The hit rate would be determined by matching each female's highest interest area with the interest area of her major. Each time the highest interest area with the highest score (i.e., high-point code) matches the interest area of an individual's major, this match is counted as a hit. There are several possible choices of chance rates by which to compare this predictive accuracy. The choice of base rate may be the proportion of all college students whose majors are classified within each interest area, the proportion of all females who major in each area, the chance of receiving the highest interest score on each scale based on the total number of possible scales, and so on.

Many of the primary studies included in this analysis did not incorporate a base rate comparison. However, one base rate that was commonly used was a chance rate of 16.67%, or about 17%, for the six RIASEC interest categories (e.g., Hughes, 1972; Latona, Harmon, & Hastings, 1987; Lattimore & Borgen, 1999; Prediger, 1998). Although this base rate comparison is better than no comparison at all, there are problematic assumptions with the use of this base rate. Namely this rate assumes equal chance of classification in any of the six interest categories. This equal chance is unlikely to be the case in reality.

Since these hit rates determine the accuracy of interest inventories in predicting occupational membership and related criterion, a more accurate base rate derivation would be the proportions of the workforce that are employed in each interest area (Su, 2018). A recent study by DeCeanne, Lewis, and Rounds (2017) determined these base rates in the U.S. employment population by merging the Bureau of Labor Statistics' (BLS) 2014 Employment Projection data with the occupational interest profiles from the Occupational Information Network (O*NET). The results of this study indicate that the employment distribution rates were not equal. Based on the distribution of over 150 million employees in the United States in 2014, the percentages of the population employed in each interest area are: 30.3% Realistic, 5.5% Investigative, 1.7% Artistic, 17.9% Social, 21.9% Enterprising, and 22.7% Conventional (DeCeanne et al., 2017). In other words, there is the highest chance of being employed in the Realistic area, and there is the lowest chance of employed in the Artistic area. These percentages can be used to evaluate the different effect size estimates for the RIASEC categories to determine whether the hit rates are higher than what would be expected by chance.

For other moderator analyses, base rate derivations are more difficult. In a meta-analytic framework, there is an added level of difficulty in determining which base rates should be used because effect sizes are combined across studies that have different underlying chance rates. Most moderator analyses aggregate across inventories, scale types (i.e., occupational scales, RIASEC interest scales, etc.), and other study characteristics, so it would be difficult to determine a base frequency by which to compare the predictive accuracy of the inventories. Ultimately, one goal of the present study is to re-introduce base rates into the conversation of predictive accuracy and urge further studies to take base rates into account for their specific samples, inventories, and methods of determining hit rates.

The Present Study

The present study attempts to provide a quantitative summary of the criterion-related validity of interest inventories in predicting career choice. Broadly speaking, career choice includes choice of occupations, college majors or fields of study, vocational aspirations, and expressed vocational plans. I meta-analytically examine the accuracy of interest inventories in predicting these criteria both concurrently and predictively. Additionally, this study is an attempt to examine different characteristics that moderate the degrees of accuracy.

First, I examined differences in accuracies between concurrent hit rates and predictive hit rates. In line with previous meta-analyses of longitudinal vocational interest studies (Hoff, Briley, Wee, & Rounds, 2018; Low, Yoon, Roberts, & Rounds, 2005), I expect the criterion assessment time to behave similarly to test-retest intervals. In this case, concurrent validity studies have no test-retest interval period, whereas predictive validity studies have a range of test-retest intervals greater than zero. Recent meta-analyses indicate that interests tend to change over time (Hoff et al., 2018; Low et al., 2005). Thus, I expect that predictive validity studies will have lower predictive accuracies than concurrent validity studies since the time between interest measurement and criterion assessment is greater, so there is more time for changes in interests, majors, and jobs.

I also included gender of the sample as a possible moderator of predictive accuracy. Most primary studies reported hit rates separately for male and female samples, indicating that it may be important to examine differences between gender groups. Meta-analytic evidence indicates that there are established gender differences in vocational interests (Su, Rounds, & Armstrong, 2009), which are also reflected in differences in employment rates for certain fields such as STEM (science, technology, engineering, and math) fields (Su & Rounds, 2015). Based on these gender differences, men and women may have different underlying base rates of working in different occupations or choosing different majors, so the predictive accuracies may also differ.

Additionally, there is a long history of debate regarding the optimal way to score interest inventories and present career suggestions to males and females (Lamb & Prediger, 1979). Sex differences in interests influenced the development of interest inventory scales, particularly the occupational scales on the Strong Interest Inventory (SII; Harmon et al., 1994). The development of occupational scales for the SII involved the use of separate norm groups for males and females. For each occupational scale, an individual may be scored according to their same-sex norms, opposite-sex norms, standard-score or combined-sex norms, or raw scores. Various studies have utilized different norming techniques in calculating hit rates by gender (Betz & Wolfe, 1981; Cairo, 1982; Dik & Hansen, 2004; Dolliver & Worthington, 1981). Studies using other inventories, such as the ACT Interest Inventory (UNIACT-R; Swaney et al., 1995), have also compared the accuracy of different norming methods (Hanson, Noeth, & Prediger, 1977; Lamb & Prediger, 1979; Prediger & Lamb, 1981). In the present meta-analysis, I examine scale norming as a moderator of predictive accuracies to derive a quantitative estimate to inform this debate.

In many instances, primary studies also reported hit rates separately for different RIASEC interest criterion samples (e.g., Gottfredson & Holland, 1975; Holland & Lutz, 1968; Mount & Muchinsky, 1978; Salomone & Slaney, 1978). Based on the different employment frequencies across the interest categories (DeCeanne et al., 2017), it is likely that there will be different hit rate accuracies between the RIASEC criterion groups. In particular, I predict that the highest accuracy rates will be for the Realistic interest category since the largest majority of employees work in occupations classified as Realistic. In a similar way, the lowest hit rate accuracy will most likely be in the Artistic category since the smallest proportion of the population work in this interest sector.

Aside from various sample characteristics, there are several other study characteristics that may moderate the predictive accuracy of interest inventories. One of these possible moderators is the interest inventory used. There were over 30 different interest inventories used in the body of literature included in the present meta-analysis, but several interest inventories were most often studied: the Strong Interest Inventory (SII, SVIB, SCII; Campbell, 1971; Harmon et al., 1994; Strong, 1981), the Self-Directed Search (SDS; Holland et al., 1994), the Vocational Preference Inventory (VPI; Holland, 1985), the ACT Interest Inventory (UNIACT; Swaney, 1995), the Campbell Interest and Skill Survey (CISS; Campbell, Hyne, & Nilsen, 1992), and the Kuder Preference Record (KPR, KOIS; Kuder, 1970).

I examined the different hit rates derived from each of these interest inventories, as well as the remainder of the interest inventories grouped into an "Other" category for the purpose of analyses. Since these interest inventories have different methods of scale construction, different theoretical orientations, different types and numbers of scales, and various other distinctions between them, it is important to consider the different inventories as a moderator. Additionally, these inventories have different reputations and frequencies of use, so examining the differential accuracies may by inventory may serve as a proxy for quality (Su et al., 2009). Relatedly, I also examined the different types of scales (i.e., occupational scales, basic interests/area scales, RIASEC interest scales, and specialty scales such as medical specialties) as a potential moderator due to the variety in specificity levels and differences in scale construction methods.

In our broad inclusion of career choice criterion, there were several different types of criterion that were utilized in the primary studies. I included studies that predicted occupational membership, college major choice, vocational aspirations, or expressed career plans. Since these specific criterion categories varied across studies, I examined criterion as a potential moderator as well.

Like the choice of criterion, researchers were faced with several methodology choices while designing these criterion-related validity studies. Two important decisions included the method used to match each criterion choice to a scale on the inventory, and the method used to determine a hit, or correct prediction. In each of these cases, there were a wide variety of choices made across the studies. I examined both the criterion-scale match method and the hit rate calculation method as possible moderators as these choices could impact the stringency and quality of conclusions drawn.

Finally, I examined publication status as a possible moderator. In our search of the literature, I located both published and unpublished studies. In particular, unpublished studies were primarily drawn from dissertations, research reports, and interest inventory manuals. I included this variable as a potential moderator to test whether higher hit rate accuracies were published in peer-review journals compared to the hit rate accuracies reported in unpublished outlets.

CHAPTER 2: METHODOLOGY

Literature Search

To identify studies for the present meta-analysis, I searched for published and unpublished studies that investigated the criterion-related validity of interest inventories in predicting career choice. In the Fall of 2015, I searched for relevant articles on the PsycINFO database, Proquest Dissertations, and Google Scholar using combinations and variations of the following search terms: *interest inventories, occupation, validity, concurrent validity, predictive validity, evidence, hit rate, predictive efficiency, and vocational interests as predictors*. I also searched through available technical manuals for interest inventories, the American College Testing (ACT) Research Reports, and relevant book reviews for possible citations or data references. Finally, the reference lists of all relevant studies were examined for possible cross-referenced studies. This process resulted in 335 possible studies after eliminating duplicate search results and titles and abstracts that were clearly unrelated (i.e. did not use an interest inventory, etc.).

Inclusion Criteria

I included studies in the meta-analysis if they met several criteria. First, at minimum, studies needed to use at least one interest inventory and needed to provide a sample size. Second, I included studies that either explicitly provided a hit rate or provided data by which to calculate the percentage of correct predictions (i.e., number of participants who received a hit and total number of participants). Studies that reported other metrics of criterion-related validity (e.g., correlations between two inventories, Tilton's percent overlap, etc.) were excluded. Third, studies needed to use interest inventory scores as predictors of a career choice criterion, including occupational membership, occupational group or industry, college major choice or field of study, vocational aspirations or preferences, or expressed occupational plans. Studies that used predictors other than interest inventories (e.g., expressed career choice, skills, values, etc.) were excluded. Additionally, studies that predicted criterion other than career choice were excluded. Finally, studies needed to be published in English.

I conducted the literature search and independently retrieved each study. A total of 151 studies met the inclusion criteria. Among these studies, there were several instances in which the same sample of participants was used multiple times. When multiple studies analyzed the same sample of participants, these studies were coded as a single "composite" study. For example, if five studies used the same sample of participants, these five were coded as a single study so they would not be counted multiple times in the meta-analysis.

Each of these composite studies contributed a single overall effect size to the meta-analysis. To attain the single effect size, I calculated a sample-size weighted effect size using the sample size from each study included in the composite. With the composites coded as a single study, there were a total of 130 unique studies with 345 unique samples (N = 309,033). These studies yielded a total of 1,965 hit rates.

Study Variables

Hit rate. The effect size in this study was the hit rate, or percentage of correct predictions. In particular, this percentage refers to the number of correct predictions made by the interest inventory out of the total number of predictions made. In cases where a primary study listed the hit percentage, this percentage was recorded directly. In cases where a primary study did not report the hit percentage, this percentage was calculated using the information provided.

Aggregation of effect sizes. In most cases, a single study reported several hit rates. For example, studies often reported a hit rate for males and a hit rate for females, in which case these would be considered two unique samples. Additionally, studies often reported several hit rates for the same sample, such as an overall female hit rate, as well as hit rates for different sub-samples of females with different types of majors. In total, there were 1,965 hit rates reported.

To address the issue of non-independence, I first aggregated hit rates at the sample level by computing a sub-sample weighted hit rate for each unique sample. For the example described above, the sub-samples of females in different majors were aggregated to one female hit rate for the sample. This aggregation process resulted in 345 sample-level hit rates.

Then, I derived a sample-size weighted composite hit rate at the level of each study. For the example described above, males and females from the same study were aggregated to one study hit rate. This aggregation process resulted in 130 independent study-level hit rates. As a sensitivity analysis of the independence assumption for the overall meta-analysis of hit rates, I meta-analyzed both the 345 sample-level and 130 study-level hit rates and compared the effect size estimates from each method. Similarly, for each moderator analysis, hit rates were first aggregated at the sample-level within each moderator category. Then, these sample hit rates were aggregated using the same weighted-average technique to derive one study-level hit rate per moderator category.

Large sample. One large primary study (Prediger, 1998) included a college-bound sample of students with a total N = 126,194. All reported results include this sample. However, due to the relatively large weight of the effect size of that study and its potential impact on results, all results were also computed without this sample. The results of the moderator analyses that were computed without this sample are reported in Table 3.

Criterion assessment time. Concurrent validity studies assessed the criterion at the same time in which the inventory was administered. Predictive validity studies administered the inventory and then assessed the criterion at a later time. For predictive validity, follow-up times ranged from one week to 25 years (M = 6.04 years, SD = 4.60 years).

Year of publication. The present analysis included studies published from 1939-2014. Year of publication was included as a continuous moderator, and a visual representation of study-level hit rates by year is presented in Figure 1.

Gender. For gender, I coded each hit rate as coming from a female, male, or mixed-gender sample. Moderator analyses of gender focused on the estimation of hit rates for males and females separately rather than including the mixed-gender hit rates to ensure all samples included in the moderator test were non-overlapping.

Scale norming. Some studies used norming methods to calculate individual scores to be used for validation purposes. Possible norming methods included: same-sex, cross-sex, standard/combined-sex, or raw scores. Studies that did not explicitly report the scoring method were not included in this moderator analysis. For single-gender samples, the norming method was coded into one of the four categories listed. For mixed-gender samples, hit rates based on female-normed and male-normed scores were excluded from the moderator analyses because these could not be classified as same-sex or cross-sex. In other words, the same-sex and cross-sex categories included single-gender samples only.

RIASEC criterion interest. For studies that used inventories with Holland's (1959; 1997) RIASEC scales (i.e., Strong, SDS, VPI, ACT Inventory, etc.) as predictors, hit rates were recorded separately for each of the six criterion groups. In most cases, these hit rates were reported for separate samples of participants, so these hit rates were first aggregated at the level of six unique samples prior to being aggregated at the study level.

Occupation/major. Some studies reported separate hit rates for specific occupations or college majors. I coded occupations and majors into RIASEC categories using the Occupational Information Network's (O*NET) Interest Profiles (Rounds et al., 2008). These coded occupations were included together in the RIASEC sample hit rate analysis.

Interest inventory. There were a variety of interest inventories included in these analyses. The most commonly used inventories included the Strong Interest Inventory (SII, SVIB, SCII; Campbell, 1971; Harmon et al., 1994; Strong, 1981), the Self-Directed Search (SDS; Holland et al., 1994), the Vocational Preference Inventory (VPI; Holland, 1985), the ACT Interest Inventory (UNIACT; Swaney, 1995), the Campbell Interest and Skill Survey (CISS; Campbell, Hyne, & Nilsen, 1992), and the Kuder Preference Record and Kuder Occupational Interest Survey (KPR, KOIS; Kuder, 1966; 1970). All other interest inventories were grouped together in an "Other" category for the meta-analyses that included interest inventory as a moderator. Some examples of inventories included in this category are the Geist Picture Inventory (Geist, 1959), the Ohio Vocational Interest Survey (OVIS; D'Costa, Winefordner, Odgers, & Koons Jr, 1970), the Medical Specialty Preference Inventory (MSPI; Zimny, 1980), and the Minnesota Vocational Interest Inventory (MVII; Clark & Campbell, 1965).

Criterion. For the present meta-analysis, I included studies with criterion related to career choice. Possible criteria included occupation, college major/field of study, vocational aspiration or preference, or expressed occupational choice. All other possible criteria were excluded (e.g., expressed career choice, skills, values, etc.).

Criterion-scale matching method. An important aspect of evaluating the criterion-related validity of interest inventories involves choosing which scale on the inventory matches or most closely resembles the criterion. There are different methods of matching an occupation or major to a scale on the inventory, and the matching method depends on the type of scale. For RIASEC scales, researchers matched the criterion according to the most appropriate RIASEC category. For occupational scales, the criterion may be either directly matched to a scale (e.g., the occupation of accountant matched to the Accountant occupational scale) or indirectly matched to an occupational scale (e.g., the occupation of business analyst matched to the Accountant scale). I coded which matching method was used in each study. The match methods were grouped into the following categories: direct match, indirect match, mixed direct and indirect matches, matching the criterion to a RIASEC scale, matching the criterion to another job family grouping, or other methods.

Inventory scale type. Hit rates were calculated for different types of inventory scales: occupational scales, basic interest or other area scales, specialty scales (e.g. medical specialties), or RIASEC scales. Hit rates from combined sets of scales were not included in the moderator analysis.

Calculation of hit rates. In each primary study, a "hit" was defined in a particular way. For analyses, these various hit rate calculations were classified into one of six methods. One way to calculate hits was based on a cut score, where each person who scored above the cut score on their matched scale was considered a hit. Typically, studies used a McArthur (1954) cut score of a standard score of 40 or above on the matched occupational scale, or the equivalent standard scores for other types of scales. A second hit rate calculation method was based on RIASEC high-point codes, where a hit was recorded when an individual's interest high-point code matched the high-point code of their criterion (i.e. college major, occupation, etc.). In other words, a person received a hit when their highest inventory score was on the same RIASEC scale as the classification for their occupation, college major, or other criterion.

In a third method, some studies defined a hit as scoring highest on one's own scale relative to all other scales. For example, if an accounting major scored highest on the Accountant occupational scale, that would be considered a hit. Fourth, hits were sometimes determined based on the inclusion of the relevant scale score in a certain top proportion of scores. For example, some studies indicated that a hit was recorded if an individual's matched scale score was in the top 6% of all their scores, top five highest scores, or some other proportion-based distinction. Fifth, a smaller number of studies determined a hit by comparing an individual's score to a reference or norm group. For example, if an individual's score on their own scale was among the top 10% of the norm group scores on that scale, that would be considered a hit.

Finally, some studies used the inventory scale scores as predictor sets in a discriminant function analysis to predict the criterion group for each individual. Discriminant analysis is a methodology typically used to predict a categorical criterion outcome from a set of predictor variables, so this methodology is well-suited to predict occupational membership from a set of interest scores (Betz, 1987; Donnay & Borgen, 1996). If the discriminant analysis resulted in the prediction of the correct criterion group, that prediction was considered a hit.

Three-category distinctions. Some hit rate calculations included three categories, or levels, of hits based on stringencies of different cut scores, different high-point codes, or other distinctions. For example, a common categorical hit distinction was based on the McArthur method of scoring (McArthur, 1954). A standard score above 45 on the correct occupational scale was an "excellent" hit, a standard score of 40-44 was a "good" hit, and all scores below 40 were considered a “miss.” A related distinction was McArthur (1954) levels by rank, in which relative ranks of the scores were also considered at different levels. Another categorical distinction was based off of high-point codes: an “excellent” hit was when the criterion matched a person’s highest RIASEC score (high-point code) on the inventory, a “good” hit was when the criterion matched the person’s second highest RIASEC score, and anything else was considered a “miss.” Other categorical hits existed as well.

For these three-category distinctions, it was possible to recode the hit categories into a single-category structure to match the existing hit calculations described previously. For example, for the McArthur (1954) categories, summing the “excellent” and “good” hit categories resulted in a single hit rate that could be classified as a cut score/McArthur hit rate. For high-point codes, the “excellent” hit rate of the first-letter high-point match could be classified as a high-point match. In this way, three-category hit classifications were recoded into single hit rates.

Publication status. The publication status of the sample was coded dichotomously as published or unpublished. Unpublished studies included dissertations, unpublished data in interest inventory manuals, and research reports. To investigate the possibility of publication bias, I also examined the funnel plot in Figure 2.

Analytical Approach

The criterion-related validity of interest inventories in predicting career choice was assessed via the meta-analytic estimation of hit rates. All statistical analyses were carried out in R using the metafor package (Viechtbauer, 2010). Due to the heterogeneity of the sample, each moderator was analyzed in an independent mixed-effects meta-regression analysis. Maximum likelihood estimation was utilized for all effect size estimates. Weights for each effect size were equivalent to the inverse of the random-effects variance.

Since hit rates are represented as percentages, or proportions, I used a binomial-normal logistic meta-regression model (Stijnen, Hamza, & Özdemir, 2010). This model is a specialized case of a generalized linear random- or mixed-effects meta-analytic model. This model is most appropriate for the analysis of proportions because hit rates are based on dichotomous event counts: each person is either considered a hit (p) or not a hit ($1-p$). As such, hit rates are assumed to follow a binomial distribution (Stijnen et al., 2010).

Each proportion was transformed with a logit transformation, or log-odds of the proportion, prior to being entered into the meta-analysis. To meet the assumption of normality, all statistical tests were based on these log-odds. The logit can be calculated as $L = \log[p/(1-p)]$, where p is the proportion of correct predictions in each sample (Lipsey & Wilson, 2001; Sutton, Abrams, Jones, Sheldon, & Song, 2000). This transformation is used to extend the end-points of the proportion distribution to negative and positive infinity. By extending the ends of the scale in this manner, ceiling and floor effects are eliminated because all real numbers are included in the distribution.

In the logit formula, a hit proportion of $p = 1$ is problematic because of the resulting denominator of $1-p = 0$. Similarly, hit proportions of $p = 0$ resulted in a log[0]. This calculation results in a logit of negative infinity, which cannot be meta-analyzed. In these cases, a value of ½ was added to the problematic entry to allow the logit formula to calculate a real number value. This solution is the default treatment in the metafor package because a common solution for zero cell counts is to add a small, non-negative constant to the problematic cell (Viechtbauer, 2010).

Finally, each meta-analytic estimate was back-transformed for interpretation of the final estimates and confidence intervals. The back transformation used the formula $p = [e^L/(1+e^L)]$ (Lipsey & Wilson, 2001; Sutton et al., 2000).

CHAPTER 3: RESULTS

To derive an overall estimate of the accuracy of interest inventories in predicting career choice, I first analyzed the full set of weighted hit rates for all studies included in the analysis (k = 130, N = 309,033). The overall meta-analytic hit rate estimate at the study level is 50.3% (95% CI = [47.3, 53.4], τ = .694, I^2 = 99.52%). As a sensitivity check for the assumption of independence of effect sizes, I also derived a sub-sample weighted hit rate for every unique sample (k= 345). This analysis allowed for the inclusion of multiple unique samples per study. The sample-level meta-analytic hit rate estimate is 49.7% (95% CI = [47.2, 52.3], τ = .94, I^2 = 99.3%).

To account for the largest primary study (Prediger, 1998; N = 126,194), results were also computed without this sample. I first re-analyzed the full set of study-level hit rates (k = 129, N = 182,839). The overall meta-analytic hit rate estimate at the study level without Prediger's (1998) large sample is 50.4% (95% CI = [47.4, 53.5]), so results only changed by a magnitude of .1. Similarly, I also re-analyzed the full set of sample-level hit rates (k = 344, N = 182,839). The overall meta-analytic hit rate estimate at the sample level without Prediger's (1998) large sample is 49.8% (95% CI = [47.2, 52.3]), so once again results only changed by a magnitude of .1.

Noticeably, the overall study-level and sample-level hit rate estimates are nearly identical. The primary difference is a lower value for τ at the study level, which is the square root of the total between-study variability in effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2011). Since the study-level hit rates aggregate across samples, the hit rates naturally reflect less variation than the hit rates at the sample level. In both cases, the I^2 estimates are nearly 100%. I^2 reflects the proportion of variance in effect sizes that reflects true variation, rather than sampling error (Borenstein, Higgins, Hedges, & Rothstein, 2017). These estimates indicate that almost all of the variation in effect sizes is due to true variation. As a result, I then proceeded with a mixed-effects meta-regression model to examine various moderators that may explain the heterogeneity in effect sizes.

Sample and Study Characteristics

The results for the moderator analyses are presented in Table 1. Each moderator was examined in an independent meta-regression analysis. For each moderator, hit rates were aggregated using sample-size weighted average hit rates within each moderator category for each study. Additionally, the significance test of moderators in each case was based on the logit-transformed proportions that were entered into the meta-analysis. In some cases, the logit confidence intervals for different moderator categories are non-overlapping, indicating significant differences, but the back-transformed proportions could still have overlapping confidence intervals based on the calculations used. In other words, it is important to take the test of moderators into account in order to determine significant differences because these meta-analytic differences may not always be reflected in the back-transformed proportion results.

Time. I examined time as a potential moderator in two ways. Namely, I estimated hit rates separately for concurrent and predictive studies, and I regressed the effect sizes onto publication year. First, as shown in Table 1, concurrent studies (k = 90, N = 242,564) have a higher estimated predictive accuracy rate than predictive studies (k = 62, N = 63,924). The test of moderators is significant (QM (df =1) = 4.46, p<.05). Interest inventory scores accurately predict career choice in an estimated 54.6% (95% CI = [51.7, 57.5]) of cases when the inventory and criterion are measured at the same time. On the other hand, the accuracy rate is only 40.3% (95% CI = [36.5, 44.2]) when interests and the criterion are assessed at different time points.

For all moderators that included Prediger's (1998) large college-bound sample, I re-calculated the moderator results without this sample. The moderator results without this sample are presented in Table 3. Like the overall meta-analytic estimate, every moderator estimate only changed by a magnitude of .1 except for the concurrent versus predictive criterion assessment. In this analysis, the estimates changed more substantially when Prediger's (1998) college-bound sample was removed. As shown in Table 3, concurrent studies (k = 89, N = 116,370) still have a higher estimated predictive accuracy rate than predictive studies (k = 62, N = 63,924). The estimated hit rate for concurrent studies went from 54.6% (95% CI = [51.7, 57.5]) to 52.9% (95% CI = [49.1, 56.6]), and the estimated hit rate for predictive studies went from 40.3% (95% CI = [36.5, 44.2]) to 46.4% (95% CI = [42.0, 50.9]). Despite the relatively larger magnitude of estimated effect size changes, the test of moderators is still significant (QM (df =1) = 4.68, p<.05), so the moderator test ultimately does not change in interpretation.

In addition to concurrent versus predictive criterion assessment, I also examined publication year. The earliest year of publication in the present meta-analysis is 1939 (Dyer, 1939), and the most recent studies were published in 2014 (Burns, 2014a,b). The median publication year is 1979. Figure 1 shows a visual display of hit rates by publication year, with the size of points reflecting the sample size as a measure of precision. As shown in Figure 1, there appears to be a slight negative trend over time in the raw data. However, the test of moderators in the meta-regression analysis is not significant (QM (df =1) = 2.57, p = .11), indicating that hit rate accuracies do not significantly differ across time of publication.

Gender and norming. One of the primary sample characteristics of interest in the present meta-analysis is gender. These results are displayed in Table 1. For gender, the test of moderators is not significant (QM (df =1) = .47, p = .49). The estimated hit rate for males (k =

87, *N* = 51,851) is 51.1% (95% CI = [47.2, 55.1]), and the estimated hit rate for females (*k* = 68, *N* = 42,384) is 49.0% (95% CI = [44.6, 53.5]). Counter to our initial expectation, there is no significant difference between males and females in the accuracy of interest inventories in predicting career choice outcomes.

Related to gender, I also examined different accuracy rates based on how the inventories were norm-scored. Again, the test of moderators is not significant (*QM* (*df* =3) = 6.79, *p* = .08). As shown in Table 1, there are differences in the predictive accuracies of different types of normed scores, but the confidence intervals overlap. Based on the directional trends of the hit rate estimates, the norming methods that are most effective are same-sex norms (*k* = 27, *N* = 37,054) with a hit rate of 54.0% (95% CI = [47.2, 60.6]) and combined-sex norms (*k* = 13, *N* = 14,059) with a hit rate of 60.4% (95% CI = [50.8, 69.2]). Although the differences do not reach the level of significance, the effect size estimates convey more accuracy than cross-sex norms (*k* = 11, *N* = 4,413) with a hit rate of 43.2% (95% CI = [33.2, 53.9]) and the use of raw scores (*k* = 8, *N* = 27,454) with a hit rate of 46.0% (95% CI = [34.3, 58.2]).

Interest categories and base rates. There are some differences in the hit rates between the RIASEC criterion groups, and the test of moderators is significant (*QM* (*df* = 5) = 12.57, *p*<.05). As shown in Table 1, the highest hit rates appear to be for Investigative criterion groups (*k* = 62, *N* = 26,945) with a hit rate of 50.0% (95% CI = [45.1, 54.9]), Realistic criterion groups (*k* = 45, *N* = 9,391) with a hit rate of 49.0% (95% CI = [43.1, 54.9]), and Social criterion groups (*k* = 48, *N* = 17,998) with a hit rate of 45.5% (95% CI = [40.0, 51.2]). The Artistic (*k* = 42, *N* = 12,133) hit rate of 39.2% (95% CI = [33.5, 45.2]) is considerably lower than Investigative, Realistic, and Social, as I expected based on the distribution of employees in the workforce (DeCeanne et al., 2017).

Ultimately, the RIASEC hit rates describe the percent of correct predictions made by the inventories regarding the high-point code of each person's occupation, major, vocational aspiration, or expressed choice. To most accurately compare the RIASEC employment base rates to RIASEC hit rates, I examined the meta-analytic hit rate estimates specifically for the subset of studies predicting occupational membership. Using this subset, I can compare the accuracies of the interest inventories in predicting occupational high-point codes relative to the frequencies of employment in each interest category in the general population.

The results of this analysis are presented in Table 2. In addition to the RIASEC occupational hit rates, the proportions of the U.S. population employed in each interest area are presented (DeCeanne et al., 2017). In the last column, I took these base rates into account by subtracting the employment distribution base rates from the original hit rate estimates (Worthington & Dolliver, 1977). For example, the Realistic (k = 13, N = 2,930) occupational hit rate of 53.9% (95% CI = [42.4, 65.1]) has an accuracy of 23.6% above and beyond the employment base rate of 30.3%. In other words, the chance of guessing that someone works in a Realistic occupation is roughly 30% based on the employment distribution, so the interest inventories correctly predict Realistic occupational membership about 24% greater than chance. The more people that are employed in an interest area, the higher the chance rate of correctly predicting that a person works in that interest category.

Conversely, interest categories with relatively smaller proportions of the population are more difficult to predict by chance. Since Investigative (k = 29, N = 9,998) has a relatively small base rate of 5.5%, the occupational hit rate of 57.2% (95% CI = [49.7, 64.4]) has the highest hit rate of 51.7% after subtracting the base rate. Furthermore, although Artistic (k = 12, N = 848) has the lowest occupational hit rate of 40.4% (95% CI = [29.3, 52.7]), the adjusted hit rate of 38.7%

is actually higher than the adjusted hit rates for Realistic (23.6%), Social (34.7%), Enterprising (23.2%), and Conventional (25.4%) after taking into account the base rates. These comparisons provide some evidence regarding the importance of base rates in validity studies (Bokhari & Hubert, 2015; Meehl & Rosen, 1955; Schmidt, 1974).

Interest inventory and scales. Table 1 shows the results of the moderator analyses for interest inventories and for scale type. There is quite a bit of variation in the hit rate estimates by inventory, and the test of moderators is significant (*QM* (*df* = 6) = 17.87, *p*<.01). The highest hit rates are the Strong (*k* = 61, *N* = 41,061) hit rate estimate of 53.8% (95% CI = [49.5, 58.1]), the Kuder Preference Record (KPR, KOIS; *k* = 7, *N* = 5,262) hit rate estimate of 56.3% (95% CI = [43.6, 68.1]), and the Campbell Interest and Skill Survey (CISS; *k* = 5, *N* = 566) hit rate estimate of 64.2% (95% CI = [49.2, 76.9]). All of these estimates are significantly higher than the ACT Interest Inventory (UNIACT; *k* = 17, *N* = 199,656) hit rate estimate of 39.4% (95% CI = [32.3, 47.1]). The pattern of inventory hit rates may reflect the specificity of the scale types such that more narrow occupational scales match the bandwidth of occupations and produce higher hit rates than those of the more general RIASEC interest scales.

Indeed, the results of the scale analysis support this pattern. Again, the test of moderators is significant (*QM* (*df* = 4) = 18.29, *p*<.01). As shown in Table 1, occupational scales (*k* = 53, *N* = 19,686) have the highest hit rate estimate of 57.5% (95% CI = [52.4, 62.4]), followed by basic interest scales (*k* = 27, *N* = 42,532) with a hit rate estimate of 55.9% (95% CI = [48.9, 62.7]). Conversely, RIASEC interest scales (*k* = 64, *N* = 250,818) have the lowest estimated hit rate of 43.8% (95% CI = [39.3, 48.4]). In general, these results support the assertion that the more specific the scale, the higher the predictive accuracy.

Criterion. In the investigation of career choice, several possible criterion categories were predicted in the primary studies. These criterion categories included occupation, college major, vocational aspiration, and expressed vocational plan. As shown in Table 1, the hit rate estimates for vocational aspiration (k = 28, N = 148,132) at 45.4% (95% CI = [39.3, 51.7]) and expressed plan (k = 21, N = 50,339) at 42.5% (95% CI = [35.7, 49.7]) are at similar accuracy levels. However, the hit rate estimates for occupation (k = 55, N = 51,377) at 52.8% (95% CI = [48.3, 57.3]) and major (k = 47, N = 51,992) at 51.6% (95% CI = [46.8, 56.5]) are somewhat higher than the other less concrete criterion choices. Overall, the test of moderators is significant (QM (df = 3) = 8.05, p<.05).

Hit rate methods. Researchers had several methodological choices to make in the design of their studies. One methodological choice involved how best to match the criterion to a scale on the interest inventory. For the hit rate analysis of criterion-scale match method, the test of moderators is significant (QM (df = 4) = 25.68, p<.01). Interest inventories have the highest predictive accuracy rates when the criterion is directly matched to an inventory scale (k = 35, N = 15,191), with a hit rate estimate of 59.1% (95% CI = [53.3, 64.7]) and when there is a mix of directly-matched and indirectly-matched criterion (k = 25, N = 9,435) with a hit rate estimate of 60.6% (95% CI = [53.7, 67.0]). These hit rates may be contrasted with the slightly lower predictive accuracy rate when each criterion is indirectly matched to a scale (k = 26, N = 16,154), which produces a hit rate estimate of 54.0% (95% CI = [46.9, 61.0]). However, all of these criterion-scale match methods produce higher hit rates than that of high-point matching. When the criterion is matched to a RIASEC scale (k = 62, N = 245,558), the hit rate estimate is 43.8% (95% CI = [39.6, 48.1]).

Additionally, researchers had to determine a method for defining a prediction as a "hit" in each primary study. Some methods of hit calculation were more stringent than others. Although I do not have specific base rates for comparison, there is an expectation that more stringent hit calculation methods will have lower estimated hit rates. The test of moderators for hit calculation method is significant (QM (df = 5) = 75.04, p<.01). Two of the most liberal hit calculation methods include counting a hit when a person's matched scale score is among some top proportion of all their scores (e.g., top 6% of scores) or counting a hit when a person scores higher or on their own scale than a reference or norm group scores on average. As expected, the hit rate for the top proportion of scores method (k = 14, N = 18,904) is 68.2% (95% CI = [60.1, 75.4]), and the hit rate for the reference group comparison method (k = 7, N = 2,018) is 71.1% (95% CI = [59.8, 80.2]). Both of these methods result in high predictive accuracies because it is relatively "easy" to attain a hit with these calculations compared to some of the other calculation methods.

Similarly, many studies impose a cut score for the matched scale, which is typically the McArthur (1954) recommended cut-score for occupational scales. This cut score/McArthur method (k = 41, N = 14,836) has a similarly high estimated hit rate of 62.3% (95% CI = [57.4, 67.1]). This method of hit calculation does not take into account individuals' scores on other scales, so it is possible that although some individuals are counted as a hit based on their score on a particular scale, they may have scored even higher on other scales. In this way, this hit calculation method does not account for the relative ranks of a person's inventory results, so this method is fairly liberal as well.

On the other hand, a few methods do take into account the relative ranking of scores on a person's matched scale compared to their scores on all the other scales. These methods are more stringent hit calculations because the relevant scale scores are no longer evaluated independently. Two such rank-based hit calculations are RIASEC high-point code matches and attainment of the highest score on one's own matched scale. The RIASEC high-point code method counts a hit when a person's highest inventory score is on the same RIASEC scale as their criterion, so this method considers the rank of the six RIASEC scores. For this reason, researchers have sometimes compared this high-point code hit rate to a base rate of 16.67% because there are six scale scores. The RIASEC high-point code method (k = 52, N = 214,854) has an estimated hit rate of 43.7% (95% CI = [39.3, 48.2]). Not surprisingly, this hit rate is nearly identical to the hit rate for other RIASEC moderator distinctions (i.e., RIASEC scales and RIASEC criterion-scale matching), and this hit rate is lower than the less stringent hit rate calculation methods.

For the hit calculation method of attaining the highest score on one's own scale (k = 31, N = 31,474), the estimated hit rate is 40.7% (95% CI = [35.2, 46.5]). Not surprisingly, this hit rate is the lowest of all hit calculation methods because it imposes the most strict qualifications for counting a prediction as a hit. The last hit calculation method is discriminant analysis, which predicts an individual's criterion group using each of their inventory scale scores as predictors (Betz, 1987; Donnay & Borgen, 1996). The hit method of correct predictions from discriminant analysis (k = 21, N = 52,816) has an estimated hit rate of 42.1% (95% CI = [35.4, 49.0]). The base rate of for this method would be contingent on the number of scale scores and the number of criterion groups being predicted, so there may be some variability in the underlying hit rates in the primary studies. However, as expected, this hit rate is lower than several of the less stringent methods since discriminant analysis requires the exact prediction of one's own criterion group.

Test for publication bias. To test for possible publication bias, I examined the hit rates for published and unpublished studies separately as a potential moderator of hit rate accuracies.

Publication status was dichotomized, with unpublished studies including doctoral dissertations, research reports, and unpublished data reported in interest inventory manuals. The results of these analyses are provided in Table 1. The test of moderators is not significant (*QM* (*df* = 1) = 2.99, *p* = .08). The estimated hit rate for published studies (*k* = 102, *N* = 219,835) is 51.4% (95% CI = [48.0, 54.8]), and the hit rate for unpublished studies (*k* = 31, *N* = 85,866) is 45.2% (95% CI = [39.2, 51.3]).

Additionally, the funnel plot in Figure 2 displays the back-transformed proportion estimates on the x-axis plotted by their corresponding standard errors on the y-axis. Although there are slight asymmetries in certain spots, there does not appear to be a strong case for publication bias because there is not a large portion of asymmetry on either side of the overall estimated effect size.

CHAPTER 4: DISCUSSION

Objective One: Quantitative Hit Rate Estimate

The present meta-analysis is the first quantitative summary of the criterion-related validity of interest inventories in predicting career choice. There is a long history of research on the predictive accuracy of interest inventories. This rich body of literature dates back almost 100 years and spans across the fields of industrial/organizational psychology, counseling, education, and standardized assessment. The studies included in this meta-analysis provided validity data for 345 samples with almost 2,000 hit rates. By deriving quantitative summaries of these hit rates, I hope to shed light on the predictive accuracy of interest inventories and possible ways that predictive accuracy can be maximized.

The first objective in the present study was to quantitatively derive an estimate of the criterion-related validity of interest inventories. The present meta-analysis found that the overall accuracy rate of interest inventories in predicting career choice is 50.3%. This estimate indicates that across all inventories, criterion, and other sample and study characteristics, roughly half of all individuals are in occupations and majors that would be predicted by their interest inventory scores. This meta-analytic estimate aligns with several reviews of the predictive accuracy of interest inventories (Anastasi, 1988; Fouad, 1999). Overall, this robust hit rate estimate demonstrates a fairly substantial correspondence between measured interests and career choice outcomes. Importantly, the values of τ and I^2 indicated substantial amounts of true variance in population effect sizes. Due to the large amount of true heterogeneity, we analyzed several potential moderators in an attempt to explain the differences in effect sizes.

Objective Two: Moderators of Predictive Accuracy

The second objective in the present study was to examine various sample and study characteristics that might moderate the levels of criterion-related validity. Although the overall meta-analytic hit rate demonstrates considerable validity of interest inventories, the meta-analytic tests also demonstrated that quite a few moderators exist that explained some of the heterogeneity in population effect sizes. Namely, the predictive accuracy rates were moderated by concurrent versus predictive criterion assessment, criterion RIASEC category, interest inventory, type of scale, type of career choice criterion, criterion-scale match method, and hit rate calculation method. Noticeably, the majority of the moderators are drawn from study and methodological design choices rather than sample characteristics.

The primary sample characteristic that was examined as a moderator was gender. Surprisingly, given prior demonstrations of sex differences in interests (Su et al., 2009), gender did not moderate the predictive accuracy rates of interest inventories. In other words, it appears as though interest inventories predict career choices of males and females with essentially the same degree of precision. Despite gender differences in inventory construction and norm groups, both genders may ultimately receive viable career choice recommendations from interest inventories.

Based on the results of the study characteristics that moderated the predictive accuracy rates, some recommendations can be made regarding the most accurate ways to score and predict career choice. Although the differences in scale norming did not reach the level of significance, the hit rate effect sizes indicate that some norming methods produce more accurate results than others. Namely, interest inventory scoring that uses either same-sex norming or combined-sex norming results in more accurate predictions than raw scores or cross-sex norms. Combined-sex norms have a predicted accuracy of about 60%, which indicates that norm groups do not necessarily need to be separated by gender in the future. This is a potentially beneficial finding because combined samples have the capability to use larger sample sizes when the two gender groups are included in the same sample.

Additionally, although there are many available interest inventories, some have higher predictive accuracies than others. In general, there appears to be a pattern of higher hit rates for inventories that use more specific scale types such as occupational scales compared to broad-band RIASEC scales. However, an alternative explanation may be the theoretical orientation underlying these inventories. Occupational scales are typically derived empirically, whereas the RIASEC scale are developed using the rational/theoretical method (Burish, 1984; Holland, 1959; 1997; Strong, 1943). These different scale construction methods may lead to variations in the predictive accuracies of the inventories.

Another notable finding is the disparity in meta-analytic effect size estimates between different methods of calculating hit rates. The pattern of results indicates that more stringent hit calculations result in lower hit rates, as would be expected. This finding has implications for future research and counseling applications. By using a more liberal hit calculation, hit rates may be artificially inflated if base rates are not taken into account. For example, hit calculations using a cut score method should take into account an individual's frequency of scoring above that cut score on any scale. In other words, the base rate involves an individual's profile level. There are many documented discussions of response bias in interest inventory responding (Jackson, 1977; Prediger, 1998). Individuals who have a tendency to choose the extreme scale-points will have elevated interest profiles (Prediger, 1998), which also elevates their chances of scoring above the cut score on their matched scale. Ultimately, it would be beneficial for future studies to calculate hits using more stringent methods and to compare those hit rates to the appropriate base rates that would be expected by chance.

Assessment of publication bias. To address the possible existence of publication bias, I included publication status as a moderator, as well as including a funnel plot of study effect sizes (see Figure 2). Although there were slight differences in the hit rate estimates for published (51.4%) and unpublished studies (45.2%), the test of moderators was not significant, and the confidence intervals were overlapping. Based on these results, there does not appear to be much evidence of publication bias. Since hit rates are not typically evaluated based on statistical significance, publication bias was not necessarily expected in this set of studies. Regardless, these quantitative results suggest that similar predictive accuracies tend to be found in both published and unpublished data. The funnel plot in Figure 2 also supports the results found in the moderator analysis of publication status that publication bias does not seem to be an issue. Despite slight asymmetries, there does not appear to be a pattern of asymmetry or a large gap of missing results, so I concluded that there is no systematic evidence of publication bias.

Objective Three: Incorporate Base Rates into Predictive Accuracy

The third objective of the present study was to underscore the importance of incorporating base rates into the evaluation of predictive accuracy. In line with recommendations (Bokhari & Hubert, 2015; Meehl & Rosen, 1955; Schmidt, 1974), base frequencies or chance rates of occurrence should be taken into account when drawing conclusions about an instrument’s predictions. For instance, suppose an instrument could accurately predict whether a coin will land on heads or tails in about 40% of all coin-flip trials. In other words, the instrument will correctly predict the outcome for four out of every ten coin flips. This accuracy rate on its own may be evaluated as being fairly good, but the base rate in this case is 50% across trials. By taking the base rate into account, the instrument does not predict coin toss outcomes any better than chance (and in fact does worse than chance in this particular example).

Many of the primary studies included in the present meta-analysis did not consider base rates when drawing conclusions about inventory predictions. For those that did, the most common base rate used was 16.67% for the six RIASEC categories. However, as mentioned previously, this base rate assumes equal chance of employment in any of the interest categories. For prediction of employment in Holland's (1997) RIASEC categories, I argue that the most accurate base rates are population rates of employment within each interest category.

In the present case, I utilize the U.S. employment distribution rates derived by DeCeanne et al. (2017) as base rates by which to compare the accuracy of inventories predicting RIASEC occupational membership criterion. In doing so, I found that although the hit rate for Artistic occupations (39.2%) was the lowest on its own, this hit rate reflects the relatively small proportion of employees that work in that interest area. Relative to its small base rate of employment, interest inventories preform about 38% better than would be expected by chance. Similarly, the hit rate for Realistic occupations (49.0%) was one of the highest on its own, but this hit rate also reflects the high rate of employment in this interest category. After taking the base rate into account, interest inventories predict occupational membership in Realistic occupations about 24% better than chance. Ultimately, base rates help to tell the full story of how accurate the interest inventories truly are in these cases.

Limitations and Future Directions

The present study is the most comprehensive evaluation of the criterion-related validity of interest inventories in predicting career choice. The meta-analytic hit rate estimates represent quantitative summaries of studies over an extensive timespan (1939-2014) incorporating a large total sample of individuals (*N* = 309,033). One of the primary strengths of this study is the incorporation of both concurrent and predictive studies to derive an estimate of the full criterion-related validity spectrum.

However, there are several limitations to the present work. First, many existing meta-analyses of vocational interests examine the effect size metrics of correlations or mean differences (e.g., Hoff et al., 2018; Nye et al., 2012; Su et al., 2009; Van Iddekinge et al., 2011). On the other hand, the present meta-analysis is a quantitative summary of hit rates. Correlations and mean differences have less subjectivity of measurement than hit rates because these metrics are statistically-based calculations. On the other hand, hit rates represent the percent of correct predictions, but a "correct" prediction may be determined in different ways across different studies. Indeed, I include this degree of variation as a possible moderator through the examination of different hit calculation methods, and there are differences in the predictive accuracies across multiple methods of defining a hit. Essentially this moderator analysis indicates that not all hit rates are created equal, so this imposes a limitation on the precision of inferences that can be drawn. Nonetheless, a quantitative summary of these hit rates is still a necessary and important contribution to the existing literature on interest inventory validity.

Another limitation is the lack of a base rate metric by which to directly compare the majority of the moderator analyses included in this study. Although the present study explicitly draws attention to the importance of including base rates in evaluations of predictive accuracy, I do not have a precise method of deriving base rates for most analyses due to the aggregation of effect sizes across various study characteristics. I urge the inclusion of base rates in future predictive studies and encourage the use of accurate base rates that match the frequencies of the criterion being predicted. The interpretations of the meta-analytic hit rate results found in the current study must be made without these chance rate comparisons in most cases.

The future of interest inventory research should focus on the continued improvement of measurement and test development, and the present results can help to inform decisions such as scoring, specificity of scale composition, and the use of combined-sex norm groups. Future research should also direct attention to the different hit rates between interest inventories. In particular, research should attempt to differentiate between the competing explanations of scale specificity and theoretical orientation as having potential effects on predictive accuracies.

CHAPTER 5: CONCLUSION

The present meta-analysis demonstrated that interest inventories have a substantial level of criterion-related validity in predicting career choice. Specifically, interest inventories accurately predict individuals' career choices about half the time. This prediction rate indicates that measured interests are important predictors of both educational and occupational membership. Importantly, the hit rates are moderated by several characteristics, including the amount of time between inventory administration and criterion measurement, the interest category of the criterion, the particular interest inventory used, the type of scale on the inventory, the particular type of career choice criterion, the method used to match criterion to a scale, and the hit rate calculation method. Overall, these results shed light on the predictive accuracy of interest inventories in predicting career choice, as well as the different conditions under which accuracy rates may be expected to decrease or increase.

TABLES

Table 1. Hit Rate Estimates for Sample and Study-Design Characteristics

Variables	*k*	*n*	Hit Rate	95% CI	τ	I^2
Validity Type					.701	99.4
Concurrent	90	242,564*	54.6%	[51.7, 57.5]		
Predictive	62	63,924	40.3%	[36.5, 44.2]		
Gender					.733	98.54
Males	87	51,851	51.1%	[47.2, 55.1]		
Females	68	42,384	49.0%	[44.6, 53.5]		
Scale Norming					.702	99.19
Same-sex	27	37,054	54.0%	[47.2, 60.6]		
Cross-sex	11	4,413	43.2%	[33.2, 53.9]		
Combined-sex/Standard	13	14,059	60.4%	[50.8, 69.2]		
Raw scores	8	27,454	46.0%	[34.3, 58.2]		
Interest category					.765	97.35
Realistic	45	9,391	49.0%	[43.1, 54.9]		
Investigative	62	26,945	50.0%	[45.1, 54.9]		
Artistic	42	12,133	39.2%	[33.5, 45.2]		
Social	48	17,998	45.5%	[40.0, 51.2]		
Enterprising	49	11,937	40.2%	[34.9, 45.7]		
Conventional	40	9,135	43.2%	[37.1, 49.4]		
Inventory					.671	99.31
Strong	61	41,016	53.8%	[49.5, 58.1]		
SDS	14	7,915	43.3%	[34.7, 52.3]		
UNIACT	17	199,656	39.4%	[32.3, 47.1]		
Kuder	7	5,262	56.3%	[43.6, 68.1]		
VPI	14	19,474	46.2%	[37.6, 55.0]		
Campbell	5	556	64.2%	[49.2, 76.9]		
Other	24	25,339	52.4%	[45.5, 59.2]		
Scale Type					.736	99.48
Occupational Scales	53	19,686	57.5%	[52.4, 62.4]		
Basic Interests/Areas	27	42,532	55.9%	[48.9, 62.7]		
Specialty scales	6	3,843	45.5%	[31.3, 60.4]		
RIASEC	64	250,818	43.8%	[39.3, 48.4]		

Table 1 cont'd.

Variables	*k*	*n*	Hit Rate	95% CI	τ	I^2
Criterion					.663	99.07
Occupation	55	51,377	52.8%	[48.3, 57.3]		
Major	47	51,992	51.6%	[46.8, 56.5]		
Aspiration	28	148,132	45.4%	[39.3, 51.7]		
Expressed plan	21	50,339	42.5%	[35.7, 49.7]		
Criterion-Scale Match					.690	99.35
Direct match	35	15,191	59.1%	[53.3, 64.7]		
Indirect match	26	16,154	54.0%	[46.9, 61.0]		
Mixed (McArthur)	25	9,435	60.6%	[53.7, 67.0]		
Matched by RIASEC	62	245,558	43.8%	[39.6, 48.1]		
Matched by job family	10	13,843	49.4%	[38.7, 60.2]		
Hit Calculation					.651	99.27
Highest score	31	31,474	40.7%	[35.2, 46.5]		
High-point codes	52	214,854	43.7%	[39.3, 48.2]		
Top proportion of scores	14	18,904	68.2%	[60.1, 75.4]		
Cut score/McArthur	41	14,836	62.3%	[57.4, 67.1]		
Vs. reference group	7	2,018	71.1%	[59.8, 80.2]		
Discriminant analysis	21	52,816	42.1%	[35.4, 49.0]		
Publication Status					.690	99.42
Published	102	219,835	51.4%	[48.0, 54.8]		
Unpublished	31	85,866	45.2%	[39.2, 51.3]		

Note: Total *k*= 130; total N= 309,033; SDS= Self Directed Search, UNIACT = Unisex Edition of the ACT Inventory, Kuder = Kuder Preference Record, VPI = Vocational Preference Inventory, Campbell = Campbell Interest and Skill Survey.

*One large primary study (Prediger, 1998) included a college-bound sample of students with a total n= 126,194. Reported results include this sample, but all results were also computed without this sample. The results of the moderator analyses that were computed without this sample are reported in Table 3. The only moderator analysis with significant levels of change was for the analysis by Validity Type.

Table 2. Comparison of RIASEC Occupational Hit Rates and Base Rates of Employment in Each RIASEC Category

Interest Category	*k*	*n*	Hit Rate	95% CI	Base Rate Employment	Hit Rate – Base Rate
Realistic	13	2,930	53.9%	[42.4, 65.1]	30.3%	23.6%
Investigative	29	9,998	57.2%	[49.7, 64.4]	5.5%	51.7%
Artistic	12	848	40.4%	[29.3, 52.7]	1.7%	38.7%
Social	15	4,079	52.6%	[42.0, 63.1]	17.9%	34.7%
Enterprising	17	4,034	45.1%	[35.4, 55.1]	21.9%	23.2%
Conventional	13	1,530	48.1%	[36.7, 59.7]	22.7%	25.4%

Note: Base rate employment percentages are drawn from DeCeanne et al. (2017) using employment distributions in the U.S. workforce in 2014.

Table 3. Moderator Analyses Without Prediger (1998) College-Bound Sample

Variables	*k*	*n*	Hit Rate	95% CI	τ	I^2
Validity Type					.702	99.15
Concurrent	89	116,370	52.9%	[49.1, 56.6]		
Predictive	62	63,924	46.4%	[42.0, 50.9]		
Inventory					.674	99.04
Strong	61	41,016	53.8%	[49.5, 58.1]		
SDS	14	7,915	43.3%	[34.6, 52.3]		
UNIACT	17	73,462	39.5%	[32.1, 47.4]		
Kuder	7	5,262	56.3%	[43.6, 68.2]		
VPI	14	19,474	46.2%	[37.6, 55.1]		
Campbell	5	556	64.2%	[49.1, 77.0]		
Other	24	25,339	52.4%	[45.5, 59.2]		
Scale Type					.739	99.27
Occupational Scales	53	19,686	57.5%	[52.4, 62.4]		
Basic Interests/Areas	27	42,532	55.9%	[48.9, 62.7]		
Specialty scales	6	3,843	45.5%	[31.3, 60.4]		
RIASEC	63	124,624	43.9%	[39.4, 48.5]		
Criterion					.665	98.94
Occupation	55	51,377	52.8%	[48.2, 57.3]		
Major	47	51,992	51.6%	[46.8, 56.5]		
Aspiration	27	21,938	45.7%	[39.4, 52.1]		
Expressed plan	21	50,339	42.5%	[35.7, 49.7]		
Criterion-Scale Match					.693	99.05
Direct match	35	15,191	59.1%	[53.3, 64.7]		
Indirect match	26	16,154	54.1%	[46.9, 61.1]		
Mixed (McArthur)	25	9,435	60.6%	[53.7, 67.1]		
Matched by RIASEC	61	119,364	43.9%	[39.6, 48.3]		
Matched by job family	10	13,843	49.4%	[38.7, 60.2]		

Table 3 cont'd.

Variables	*k*	*n*	Hit Rate	95% CI	τ	I^2
Hit Calculation					.653	99.00
Highest score	31	31,474	40.7%	[35.2, 46.5]		
High-point codes	51	88,660	43.8%	[39.4, 48.3]		
Top proportion of scores	14	18,904	68.2%	[60.1, 75.4]		
Cut score/McArthur	41	14,836	62.3%	[57.3, 67.1]		
Vs. reference group	7	2,018	71.1%	[59.8, 80.2]		
Discriminant analysis	21	52,816	42.1%	[35.4, 49.1]		
Publication Status					.691	99.21
Published	101	93,641	51.5%	[48.1, 55.0]		
Unpublished	31	85,866	45.2%	[39.2, 51.3]		

FIGURES

Figure 1. Hit Rates by Publication Year

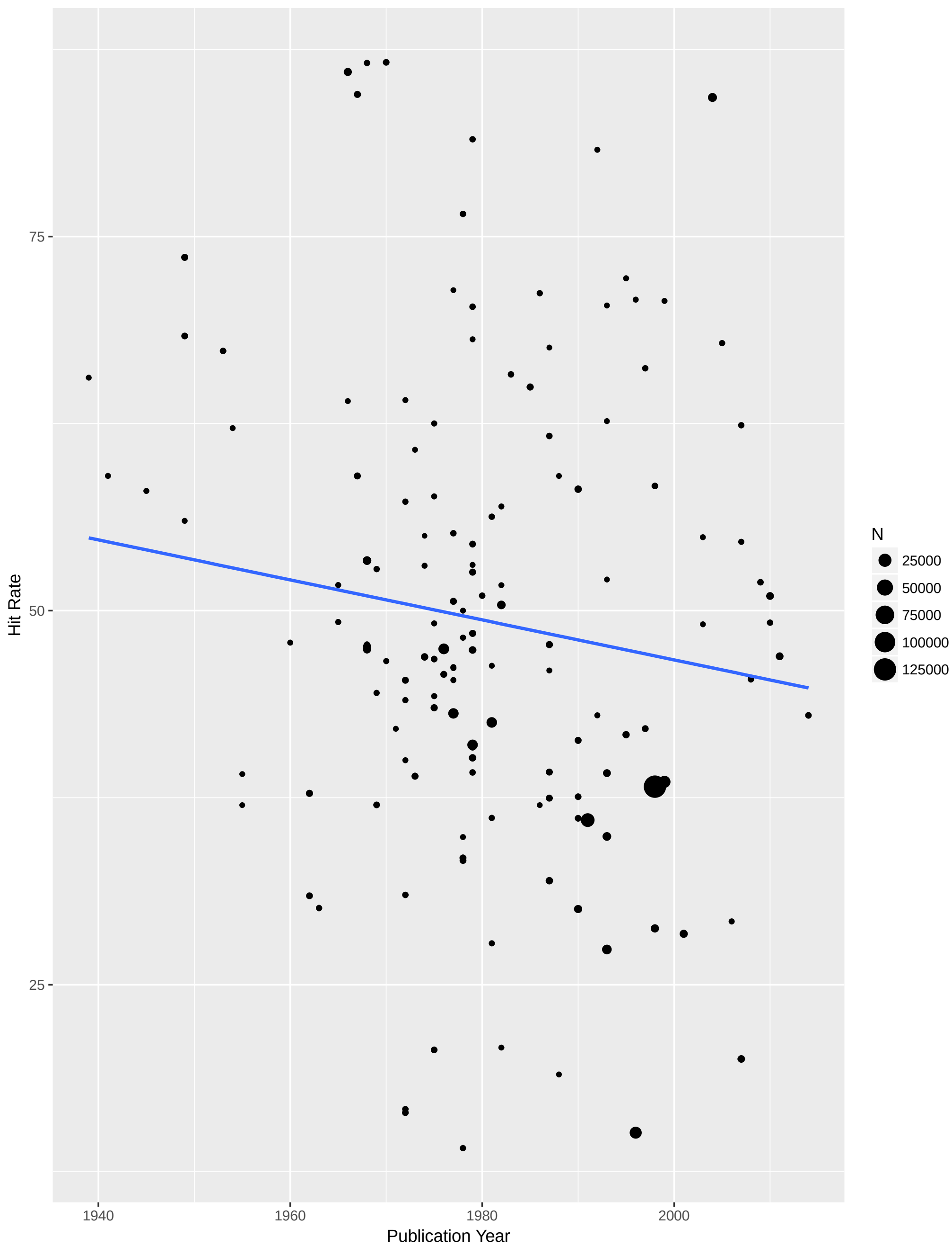
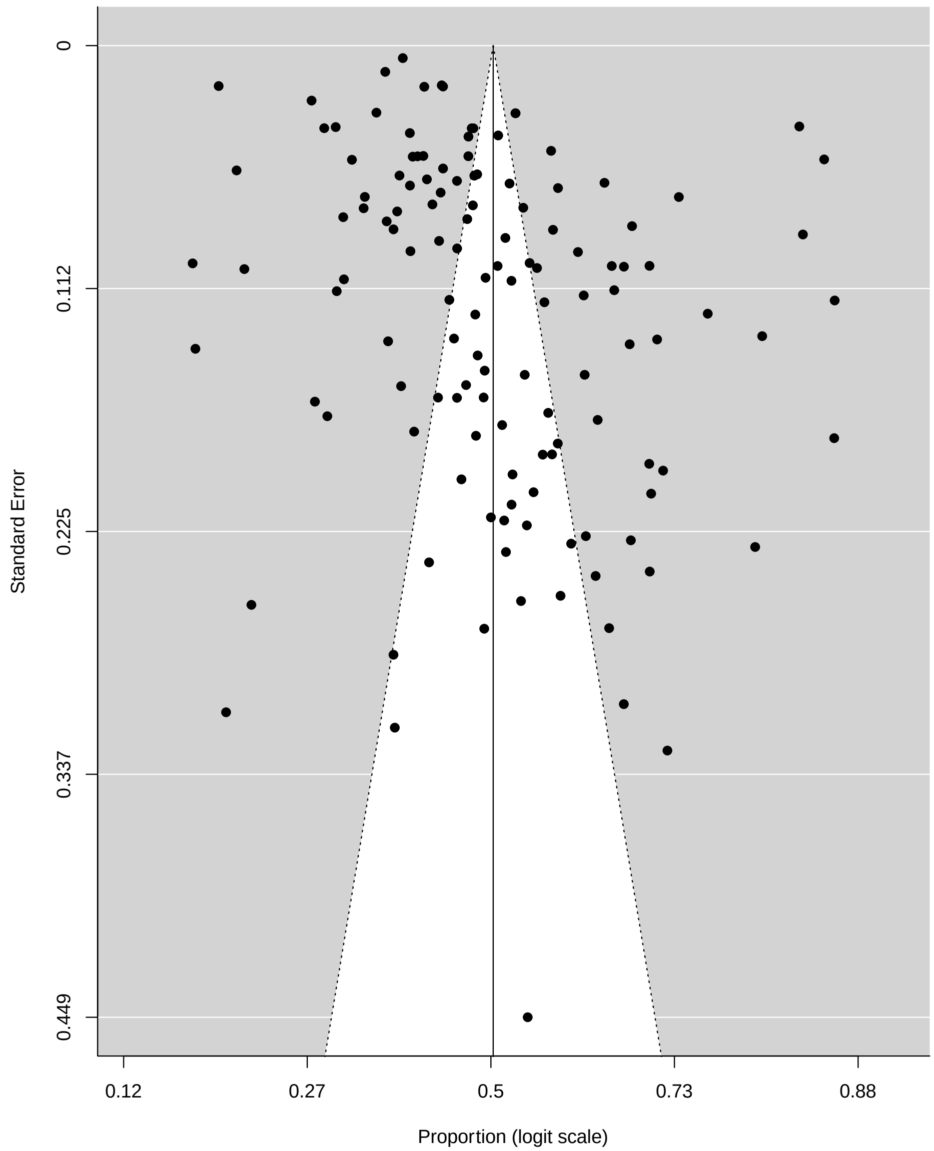


Figure 2. Funnel Plot of Study-Level Hit Rate Estimates



REFERENCES

ACT. (2009). ACT Interest Inventory Technical Manual.

Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.

*Bartling, H. C., & Hood, A. B. (1981). An 11-year follow-up of measured interest and vocational choice. *Journal of Counseling Psychology*, *28*(1), 27.

*Berdie, R. F. (1965). Strong Vocational Interest Blank scores of high school seniors and their later occupational entry: II. *Journal of Applied Psychology*, *49*(3), 188.

Betz, N. E. (1987). Use of discriminant analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 393.

*Betz, N. E., & Taylor, K. M. (1982). Concurrent validity of the Strong-Campbell Interest Inventory for graduate students in counseling. *Journal of Counseling Psychology*, *29*(6), 626.

*Betz, N. E., & Wolfe, L. K. (1981). Comparison of the utility of two approaches to sex-balanced interest scales for college women. *Journal of Vocational Behavior*, *19*(1), 61-77.

Bokhari, E., & Hubert, L. (2015). A new condition for assessing the clinical efficiency of a diagnostic test. *Psychological assessment*, *27*(3), 745.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods*, *8*(1), 5-18.

*Borgen, F. H. (1972). Predicting career choices of able college men from Occupational and Basic Interest scales of the Strong Vocational Interest Blank. *Journal of Counseling Psychology*, *19*(3), 202.

*Borgen, F. H., & Harper, G. T. (1973). Predictive validity of measured vocational interests with black and white college men. *Measurement and Evaluation in Guidance*, *6*(1), 19-27.

*Borgen, F. H., & Helms, J. E. (1975). Validity generalization of the men's form of the Strong Vocational Interest Blank with academically able women. *Journal of Counseling Psychology*, *22*(3), 210.

*Borgen, F. H., & Seling, M. J. (1978). Expressed and inventoried interests revisited: Perspicacity in the person. *Journal of Counseling Psychology*, *25*(6), 536.

*Brandt, J. E., & Hood, A. B. (1968). Effect of personality adjustment on the predictive validity of the Strong Vocational Interest Blank. *Journal of Counseling Psychology*, *15*(6), 547.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, *39*(3), 214.

*Burns, S. T. (2014a). Gender Differences in the Validity of Career Interest Inventories. *Psychology*, *5*(08), 785.

*Burns, S. T. (2014b). Validity of person matching in vocational interest inventories. *The Career Development Quarterly*, *62*(2), 114-127.

*Cairo, P. C. (1982). Measured interests versus expressed interests as predictors of long-term occupational membership. *Journal of Vocational Behavior*, *20*(3), 343-353.

*Campbell, D. P. (1966). Occupations ten years later of high school seniors with high scores on the SVIB life insurance salesman scale. *Journal of Applied Psychology*, *50*(5), 369.

Campbell, D. P. (1971). *Handbook for the strong vocational interest blank*. Stanford University Press.

Campbell, D. P., Hyne, S. A., & Nilsen, D. L. (1992). *Manual for the Campbell interest and skill survey: CISS*. National Computer Systems.

*Carter, H. D., Von Taylor, K. F., & Canning, L. B. (1941). Vocational choices and interest test scores of high school students. *The Journal of Psychology*, *11*(2), 297-306.

Clark, K. E. (1961). *Vocational interests of nonprofessional men*. Minneapolis: University of Minnesota Press.

Clark, K. E., & Campbell, D. P. (1965). *Minnesota Vocational Interest Inventory*. Psychological Corporation.

*Cooley, W. W. (1967). Interactions among interests, abilities, and career plans. *Journal of Applied Psychology*, *51*(5p2), 1.

Crites, J.O. (1984). Instruments for assessing career development. In N.C. Gysbers (Ed.), *Designing careers* (pp. 248-274). San Francisco: Jossey-Bass.

D'Costa, A. G., Winefordner, D. W., Odgers, J. G., & Koons Jr, P. B. (1970). Ohio vocational interest survey manual for interpreting. *New York, NY: Harcourt Brace Jovanovich*.

DeCeanne, A., Lewis, P., & Rounds, J. (2017). *A RIASEC Snapshot of the Modern U.S. Workforce*. Poster presentation at the 32nd Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

*Dik, B. J., & Hansen, J. I. C. (2004). Development and validation of discriminant functions for the Strong Interest Inventory. *Journal of Vocational Behavior*, *64*(1), 182-197.

*Dolliver, R. H. (1975). Concurrent prediction from the Strong Vocational Interest Blank. *Journal of Counseling Psychology*, *22*(3), 199.

*Dolliver, R. H., Irvin, J. A., & Bigley, S. S. (1972). Twelve-year follow-up of the Strong Vocational Interest Blank. *Journal of Counseling Psychology*, *19*(3), 212.

*Dolliver, R. H., & Will, J. A. (1977). Ten-year follow-up of the Tyler Vocational Card Sort and the Strong Vocational Interest Blank. *Journal of Counseling Psychology*, *24*(1), 48.

*Dolliver, R. H., & Worthington, E. L. (1981). Concurrent validity of other-sex and same-sex twin Strong-Campbell Interest Inventory occupational scales. *Journal of Counseling Psychology*, *28*(2), 126.

*Donnay, D. A., & Borgen, F. H. (1996). Validity, structure, and content of the 1994 Strong Interest Inventory. *Journal of Counseling Psychology*, *43*(3), 275.

*Donnay, D. A., & Borgen, F. H. (1999). The incremental validity of vocational self-efficacy: An examination of interest, self-efficacy, and occupation. *Journal of Counseling Psychology*, *46*(4), 432.

Donnay, D. A., Morris, M. L., Schaubhut, N. A., & Thompson, R. C. (2004). *Strong: Interest Inventory*. Consulting Psychologists Press Incorporated.

*Doty, M. S., & Betz, N. E. (1979). Comparison of the concurrent validity of Holland's theory for men and women in an enterprising occupation. *Journal of Vocational Behavior*, *15*(2), 207-216.

*Dyer, D. T. (1939). The relation between vocational interests of men in college and their subsequent occupational histories for ten years. *Journal of Applied Psychology*, *23*(2), 280.

*Enright, J. B., & Pinneau, S. R. (1955). Predictive value of subjective choice of occupation and of the Strong Vocational Interest Blank over fifteen years. *American Psychologist*, *10*, 424-425.

*Erickson, M. L. (1986). *Concurrent validity of the Strong-Campbell interest inventory for Seventh-day Adventist seminarians and ministers*. (Doctoral dissertation, Andrews University).

Fouad, N. A. (1999). Validity evidence for interest inventories. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning, measurement, and counseling use* (pp. 193-209). Palo Alto, CA, US: Davies-Black Publishing.

*Fouad, N. A., & Hansen, J. I. C. (1987). Cross-cultural predictive accuracy of the Strong-Campbell Interest Inventory. *Measurement and Evaluation in Counseling and Development*, *20*(1), 3-10.

*Gade, E. M., & Soliah, D. (1975). Vocational Preference Inventory high point codes versus expressed choices as predictors of college major and career entry. *Journal of Counseling Psychology*, *22*(2), 117.

*Gasser, C. E., Larson, L. M., & Borgen, F. H. (2007). Concurrent validity of the 2005 Strong Interest Inventory: An examination of gender and major field of study. *Journal of Career Assessment*, *15*(1), 23-43.

Geist, H. (1959). The Geist picture interest inventory: General form: Male. *Psychological Reports*, *5*(2), 413-438.

*Glavin, K. W., Richard, G. V., & Porfeli, E. J. (2009). Predictive validity of the medical specialty preference inventory. *Journal of Vocational Behavior*, *74*(1), 128-133.

*Glavin, K. W., & Savickas, M. L. (2011). Interpreting Self-Directed Search profiles: Validity of the “Rule of Eight”. *Journal of Vocational Behavior*, *79*(2), 414-418.

*Gottfredson, G. D., & Holland, J. L. (1975). Vocational choices of men and women: A comparison of predictors from the Self-Directed Search. *Journal of Counseling Psychology*, *22*(1), 28.

*Gough, H. G. (1975). Strong Vocational Interest Blank profiles of women in law, mathematics, medicine, and psychology. *Psychological Reports*, *37*(1), 127-134.

Hansen, J. I. C. (1994). The measurement of vocational interests. *Personnel selection and classification*, 293-316.

*Hansen, J. I. C., & Dik, B. J. (2005). Evidence of 12-year predictive and concurrent validity for SII Occupational Scale scores. *Journal of Vocational Behavior*, *67*(3), 365-378.

*Hansen, J. I. C., & Lee, W. V. (2007). Evidence of concurrent validity of SII scores for Asian American college students. *Journal of Career Assessment*, *15*(1), 44-54.

*Hansen, J. I. C., & Leuty, M. E. (2007). Evidence of validity for the Skill scale scores of the Campbell Interest and Skill Survey. *Journal of Vocational Behavior, 71*(1), 23-44.

*Hansen, J. I. C., & Neuman, J. L. (1999). Evidence of concurrent prediction of the Campbell Interest and Skill Survey (CISS) for college major selection. *Journal of Career Assessment*, *7*(3), 239-247.

*Hansen, J. I. C., & Sackett, S. A. (1993). Agreement between college major and vocational interests for female athlete and non-athlete college students. *Journal of Vocational Behavior*, *43*(3), 298-309.

*Hansen, J. I. C., & Swanson, J. L. (1983). Stability of interests and the predictive and concurrent validity of the 1981 Strong-Campbell Interest Inventory for college majors. *Journal of Counseling Psychology, 30*(2), 194.

*Hansen, J. I. C., & Tan, R. N. (1992). Concurrent validity of the 1985 Strong Interest Inventory for college major selection. *Measurement and Evaluation in Counseling and Development*, *25*(2), 53-57.

*Hanson, G. R. (1974). Assessing the career interests of college youth: Summary of research and applications. *ACT Research Reports, 67*, 76.

*Hanson, G. R., Noeth, R. J., & Prediger, D. J. (1977). Validity of diverse procedures for reporting interest scores: An analysis of longitudinal data. *Journal of Counseling Psychology*, *24*(6), 487.

*Hanson, G. R., & Prediger, D. J. (1973). The vocational interests of students in career-oriented educational programs. *Monograph*, *11*.

*Harmon, L. W. (1969). Predictive power over ten years of measured social service and scientific interests among college women. *Journal of Applied Psychology*, *53*(3p1), 193.

Harmon, L. W., DeWitt, D. W., Campbell, D. P., & Hansen, J. I. C. (1994). *Strong interest inventory: Applications and technical guide: form T317 of the Strong vocational interest blanks*. Stanford University Press.

*Harrington, T. F. (2006). A 20-year follow-up of the Harrington-O'Shea career decision-making system. *Measurement and Evaluation in Counseling and Development*, *38*(4), 198-202.

*Haviland, M. G., & Hansen, J. I. C. (1987). Criterion validity of the Strong-Campbell Interest Inventory for American Indian college students. *Measurement and Evaluation in Counseling and Development*, *19*(4), 196-201.

*Helms, J. C. (1979). Freshmen SDS data on entrance at a large public university. *Unpublished manuscript, University of Maryland, Baltimore*.

*Henry, P., & Bardo, H. R. (1987). Expressed occupational choice of nontraditional premedical students as measured by the Self-Directed Search: An investigation of Holland's theory. *Psychological Reports*, *60*(2), 575-581.

Hoff, K. A., Briley, D. A., Wee, C. J., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychological bulletin*, *144*(4), 426.

*Holcomb, W. R., & Anderson, W. P. (1978). Expressed and inventoried vocational interests as predictors of college graduation and vocational choice. *Journal of Vocational Behavior*, *12*(3), 290-296.

Holland, J. L. (1959). A theory of vocational choice. *Journal of counseling psychology*, *6*(1), 35.

*Holland, J. L. (1962). Some explorations of a theory of vocational choice: I. One-and two-year longitudinal studies. *Psychological Monographs: General and Applied*, *76*(26), 1.

*Holland, J. L. (1963). Explorations of a theory of vocational choice and achievement: II. A four-year prediction study. *Psychological Reports*, *12*(2), 547-594.

*Holland, J. L. (1968). Explorations of a theory of vocational choice: VI. A longitudinal study using a sample of typical college students. *Journal of Applied Psychology*, *52*(1p2), 1.

*Holland, J. L. (1972). *Professional manual for the Self Directed Search: A guide to educational and vocational planning*. Consulting Psychologists Press.

Holland, J. L. (1973). *Making vocational choices: A theory of careers* (Vol. 37). Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. L. (1985). *Vocational preference inventory*. Consulting Psychologists Press.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Psychological Assessment Resources.

*Holland, J. L., Gottfredson, G. D., & Baker, H. G. (1990). Validity of vocational aspirations and interest inventories: Extended, replicated, and reinterpreted. *Journal of Counseling Psychology*, *37*(3), 337.

*Holland, J. L., & Lutz, S. W. (1968). The predictive value of a student's choice of vocation. *The Personnel and Guidance Journal*, *46*(5), 428-434.

Holland, J. L., Powell, A. B., & Fritzsche, B. A. (1994). *The self-directed search (SDS)*. Odessa, FL: Psychological Assessment Resources.

*Hollifield, J. H. (1974). An examination of the validity of the Self-Directed Search for writers. *Measurement & Evaluation in Guidance*, *6*(4), 247.

*Holmes, T. A. (1996). *An evaluation of the concurrent validity of the 1994 Strong Interest Inventory with African American college students* (Doctoral dissertation, Pennsylvania State University).

*Hughes Jr., H. M. (1972). Vocational choice, level, and consistency: An investigation of Holland's theory on an employed sample. *Journal of Vocational Behavior*, *2*(4), 377-388.

*Isaacs, J. J., Borgen, F. H., Donnay, D. A., & Hansen, T. A. (1997). *Self-efficacy and interests: Relationships of Holland themes to college major*. Poster presented at the Annual Meeting of the American Psychological Association, Chicago, IL.

*Knapp, R. R., Knapp, L., & Knapp-Lee, L. (1985). Occupational interest measurement and subsequent career decisions: A predictive follow-up study of the COPSystem Interest Inventory. *Journal of Counseling Psychology*, *32*(3), 348.

*Knapp, R. R., Knapp, L., & Michael, W. B. (1979). The relationship of clustered interest measures and declared college major: Concurrent validity of the COPSystem Interest Inventory. *Educational and Psychological Measurement*, *39*(4), 939-945.

*Kriedt, P. H. (1949). Vocational interests of psychologists. *Journal of Applied Psychology*, *33*(5), 482.

*Kuder, G. F. (1966). *Manual, Occupational Interest Survey, Form DD.* Chicago: Science Research Associates.

Kuder, G. F. (1970). *Kuder preference record-Vocational*. Science Research Associates.

Kuder, F. G. (1977). *Activity interests and occupational choice*. Chicago, IL: Science Research Associates.

*Kunce, J. T., Dolliver, R. H., & Irvin, J. A. (1972). Perspectives on interpreting the validity of the SVIB-M. *Vocational Guidance Quarterly, 21*(1), 36-42.

*Laing, J., Lamb, R. R., & Prediger, D. J. (1982). An application of Strong's validity criteria to basic interest scales. *Journal of Vocational behavior*, *20*(2), 203-214.

*Lamb, R. R. (1976). Validity of the ACT Interest Inventory for minority group members. *ACT Research Reports*, *72*, 12.

*Lamb, R. R., & Prediger, D. J. (1979). Criterion-related validity of sex-restrictive and unisex interest scales: A comparison. *Journal of Vocational Behavior*, *15*(2), 231-246.

*Larson, L. M., Wu, T. F., Bailey, D. C., Borgen, F. H., & Gasser, C. E. (2010). Male and female college students' college majors: The contribution of basic vocational confidence and interests. *Journal of Career Assessment*, *18*(1), 16-33.

*Latona, J. R., Harmon, L. W., & Hastings, C. N. (1987). Criterion-related validity of the UNIACT with special emphasis on the World of Work Map. *Journal of Vocational Behavior, 30*(1), 49-60.

*Lattimore, R. R., & Borgen, F. H. (1999). Validity of the 1994 Strong Interest Inventory with racial and ethnic groups in the United States. *Journal of Counseling Psychology*, *46*(2), 185.

*Lau, A. W., & Abrahams, N. M. (1972). Predictive validity of vocational interests within nonprofessional occupations. *Journal of Applied Psychology*, *56*(2), 181.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.

Low, K. D., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: a quantitative review of longitudinal studies. *Psychological bulletin*, *131*(5), 713.

*Lunneborg, P. W. (1993). *Vocational Interest Inventory, Revised*. Western psychological services.

*Mau, W. C., Swaney, K. B., & Prediger, D. J. (1990). Mapping Career Groups on Basic Interest Dimensions. *ACT Research Reports, 90*, 7.

*McArthur, C. (1954). Long-term validity of the Strong Interest Test in two subcultures. *Journal of Applied Psychology*, *38*(5), 346.

*McArthur, C. (1955). Predictive power of pattern analysis and of job scale analysis of the Strong. *Journal of Counseling Psychology*, *2*(3), 205.

*McGowan, A. S. (1982). The predictive efficiency of Holland's SDS summary codes in terms of career choice: A four-year follow-up. *Journal of Vocational Behavior*, *20*(3), 294-303.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological bulletin*, *52*(3), 194.

*Millman, S. J. (1997). *Validation of the Career Type Inventory among high school students*. (Doctoral dissertation, Boston College).

Morris, M. L. (2016). Vocational interests in the United States: Sex, age, ethnicity, and year effects. *Journal of counseling psychology*, *63*(5), 604.

*Mount, M. K., & Muchinsky, P. M. (1978). Concurrent validation of Holland's hexagonal model with occupational workers. *Journal of Vocational Behavior*, *13*(3), 348-354.

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science*, *7*(4), 384-403.

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior*, *98*, 138-151.

*Olsen, L. L. (1996). *Concurrent validity of the 1994 Strong Interest Inventory: a comparison of criterion groups by gender*. (Doctoral dissertation, Iowa State University).

*O'Neil, J. M., & Magoon, T. M. (1977). The predictive power of Holland's Investigative personality type and consistency levels using the Self Directed Search. *Journal of Vocational Behavior*, *10*(1), 39-46.

*O'Neil, J. M., Magoon, T. M., & Tracey, T. J. (1978). Status of Holland's Investigative personality types and their consistency levels seven years later. *Journal of Counseling Psychology*, *25*(6), 530.

*O'Shea, A. J., & Harrington, T. F. (1971). Using the Strong Vocational Interest Blank and the Kuder Occupational Interest Survey, Form DD, with the same clients. *Journal of Counseling Psychology*, *18*(1), 44.

*Osipow, S. H., & Ashby, J. D. (1968). Vocational Preference Inventory high point codes and educational preferences. *The Personnel and Guidance Journal*, *47*(2), 126-129.

*Pendergrass, L. A., Hansen, J. I. C., Neuman, J. L., & Nutter, K. J. (2003). Examination of the concurrent validity of scores from the CISS for student-athlete college major selection: A brief report. *Measurement & Evaluation in Counseling & Development*, *35*(4).

*Perry, D. K., & Cannon, W. M. (1967). Vocational interests of computer programmers. *Journal of Applied Psychology*, *51*(1), 28.

*Perry, D. K., & Cannon, W. M. (1968). Vocational interests of female computer programmers. *Journal of Applied Psychology*, *52*(1p1), 31.

*Porfeli, E. J., Richard, G. V., & Savickas, M. L. (2010). Development of specialization scales for the MSPI: A comparison of empirical and inductive strategies. *Journal of Vocational Behavior*, *77*(2), 227-237.

*Prediger, D. J. (1987). Career counseling validity of the ASVAB Job Cluster Scales used in DISCOVER. *ACT Research Reports*, *87*, 2.

*Prediger, D. J. (1998). Is interest profile level relevant to career counseling?. *Journal of Counseling Psychology*, *45*(2), 204.

*Prediger, D. J., & Brandt, W. E. (1991). Project CHOICE: Validity of interest and ability measures for student choice of vocational program. *The Career Development Quarterly*, *40*(2), 132-144.

*Prediger, D. J., & Hanson, G. R. (1977). Some consequences of using raw-score reports of vocational interests. *Journal of Educational Measurement*, *14*(4), 323-333.

*Prediger, D. J., & Johnson, R. W. (1979). Alternatives to Sex-Restrictive Vocational Interest Assessment. *ACT Research Reports*, *79*.

*Prediger, D. J., & Lamb, R. R. (1979). The validity of sex-balanced and sex-restrictive vocational interest reports: A comparison. *Vocational Guidance Quarterly*, *28*(1), 16-24.

*Prediger, D. J., & Lamb, R. R. (1981). Four-year validity of Holland types for college-bound males and females. *Journal of College Student Personnel*, *22*(2), 133-40.

*Prediger, D. J., & Swaney, K. B. (1992). Career counseling validity of DISCOVER's Job Cluster scales for the Revised ASVAB Score Report. *ACT Research Reports, 92,* 2.

*Pulver, C. A., & Kelly, K. R. (2008). Incremental validity of the Myers-Briggs Type Indicator in predicting academic major selection of undecided university students. *Journal of Career Assessment*, *16*(4), 441-455.

*Rose, H. A., & Elton, C. F. (1970). Ask him or test him?. *Vocational Guidance Quarterly*, *19*(1), 28-32.

Rounds, J., Armstrong, P., Liao, H., Lewis, P., & Rivkin, D. (2008). *Second Generation Occupational Interest Profiles for the O*NET System: Summary.* Raleigh, NC: National Center for O*NET Development.

Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science*, *23*(2), 98-103.

*Saladin, S. A. (1996). *Concurrent validation of the 1994 Strong Interest Inventory and the Campbell Interest and Skill Survey.* (Doctoral dissertation, University of Idaho).

*Salomone, P. R., & Slaney, R. B. (1978). The applicability of Holland's theory to nonprofessional workers. *Journal of Vocational Behavior*, *13*(1), 63-74.

*Savickas, M. L., Brizzi, J. S., Brisbin, L. A., & Pethtel, L. L. (1988). Predictive validity of two medical specialty preference inventories. *Measurement and Evaluation in Counseling and Development*, *21*(3), 106-112.

Schmidt, F. L. (1974). Probability and utility assumptions underlying use of the Strong Vocational Interest Blank. *Journal of Applied Psychology*, *59*(4), 456.

*Silver, H. A., & Barnette Jr, W. L. (1970). Predictive and concurrent validity of the Minnesota Vocational Interest Inventory for vocational high school boys. *Journal of Applied Psychology*, *54*(5), 436.

*Slaney, R. B. (1978). Expressed and inventoried vocational interests: A comparison of instruments. *Journal of Counseling Psychology*, *25*(6), 520.

*Slaney, R. B., & Slaney, F. M. (1986). Relationship of expressed and inventoried vocational interests of female career counseling clients. *The Career Development Quarterly*, *35*(1), 24-33.

*Spokane, A. R. (1979). Occupational preference and the validity of the Strong-Campbell Interest Inventory for college men and women. *Journal of Counseling Psychology*, *26*(4), 312.

*Spokane, A. R. (1979). Validity of the Holland categories for college women and men. *Journal of College Student Personnel*, *20*(4), 335-340.

*Stahmann, R. F. (1969). Predicting graduation major field from freshman entrance data. *Journal of Counseling Psychology*, *16*(2p1), 109.

Stijnen, T., Hamza, T. H., & Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in medicine*, *29*(29), 3046-3067.

Strong Jr, E. K. (1927). Vocational guidance of executives. *Journal of Applied Psychology*, *11*(5), 331.

Strong Jr, E. K. (1943). *Vocational interests of men and women*. Stanford University Press.

*Strong Jr., E. K. (1955). *Vocational interests 18 years after college.* University of Minnesota Press, Minneapolis, MN.

Strong, E. K. (1981). *Strong-Campbell Interest Inventory of the Strong Vocational Interest Blank*. Stanford University Press.

Su, R. (2012). *The power of vocational interests and interest congruence in predicting career success* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Su, R. (2018). The Three Faces of Interests: An Integrative Review of Interest Research in Vocational, Organizational, and Educational Psychology. *Journal of Vocational Behavior*. 10.1016/j.jvb.2018.10.016.

Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in psychology*, *6*, 189.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological bulletin*, *135*(6), 859.

Sullivan, B. A., & Hansen, J. I. C. (2004). Evidence of construct validity of the interest scales on the Campbell Interest and Skill Survey. *Journal of Vocational Behavior*, *65*(2), 179-202.

Sutton, A. J., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). Methods for meta-analysis in medical research.

*Swaney, K. B., Lamb, R., Prediger, D., & American College Testing Program. (1995). Technical manual: Revised unisex edition of the ACT Interest Inventory (UNIACT). ACT.

*Tan, R. N. (1999). *Factors influencing the career aspirations of Asian American students*. (Doctoral dissertation, University of Minnesota).

*Tapp, R. R. L. (2001). *Structure and validity of vocational interest scales with US racial and ethnic groups: exploring the interests beyond RIASEC*. (Doctoral dissertation, Iowa State University).

*Touchton, J. G., & Magoon, T. M. (1977). Occupational daydreams as predictors of vocational plans of college women. *Journal of Vocational Behavior*, *10*(2), 156-166.

*Trimble, J. T. (1966). *A ten-year longitudinal follow-up study of inventoried interests of selected high school students.* (Doctoral dissertation, University of Missouri-Columbia).

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, *96*(6), 1167.

*Vansickle, T. R., & Prediger, D. J. (1990). Mapping occupations: A longitudinal study based on vocational interests. *ACT Research Reports*, *90*, 11.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48. URL: http://www.jstatsoft.org/v36/i03/

*Weingarten, K. P. (1953). *The measurement of interests in nonprofessional vocations by means of a pictorial inventory.* (Doctoral dissertation, University of Southern California).

*Werner, W. E. (1970). *A study of Holland's theory of vocational choice as it applies to vocational high school students.* (Doctoral dissertation, State University of New York at Buffalo).

*Whitton, M. C. (1975). Same-sex and cross-sex reliability and concurrent validity of Strong-Campbell Interest Inventory. *Journal of Counseling Psychology*, *22*(3), 204.

*Wiggins, J. D., & Weslander, D. (1977). Expressed vocational choices and later employment compared with Vocational Preference Inventory and Kuder Preference Record-Vocational scores. *Journal of Vocational Behavior*, *11*(2), 158-165.

*Wightwick, M. I. (1945). *Vocational interest patterns: A developmental study of a group of college women.* Teachers College Bureau of Publications, New York, NY.

*Williams, C. M. (1972). Occupational choice of male graduate students as related to values and personality: A test of Holland's theory. *Journal of Vocational Behavior*, *2*(1), 39-46.

*Worthington, E. L., & Dolliver, R. H. (1977). Validity studies of the Strong Vocational Interest nventories. *Journal of Counseling Psychology*, *24*(3), 208.

*Zimny, G. H. (1980). Predictive validity of the medical specialty preference inventory. *Medical Education*, *14*(6), 414-418.

*Zytowski, D. G. (1972). A concurrent test of accuracy-of-classification for the Strong Vocational Interest and Kuder Occupational Interest Survey. *Journal of Vocational Behavior*, *2*(3), 245-250.

*Zytowski, D. G. (1974). Predictive validity of the Kuder Preference Record, Form B, over a 25-year span. *Measurement and Evaluation in Guidance*, *7*(2), 122-129.

*Zytowski, D. G. (1976). Predictive validity of the Kuder Occupational Interest Survey: A 12-to 19-year follow-up. *Journal of Counseling Psychology*, *23*(3), 221.

*Zytowski, D. G., & Laing, J. (1978). Validity of other-gender-normed scales on the Kuder Occupational Interest Survey. *Journal of Counseling Psychology*, *25*(3), 205.